4. Hacia un universo digital de datos: el Big Data y Open Data

Towards a digital data Universe: Big Data and Open Data

Leopoldo Seijas Candelas
seijas.fhm@ceu.es
Universidad CEU-San Pablo

Resumen: La cantidad de información digital que se crea y almacena en la Tierra se duplica cada año como señala en su último estudio *El Universo Digital* realizado por la consultora IDC y patrocinado por EMC, uno de los primeros fabricantes de dispositivos de almacenamiento del mundo.

Los grandes volúmenes de datos traerán grandes oportunidades para empresas y negocios que sean capaces de gestionar adecuadamente esta inmensidad de datos. Se requerirá una eficiente gestión de los mismos y el uso de herramientas de software especiales dado que casi todos ellos serán no estructurados y las herramientas tradicionales no podrán administrar bien esta inmensidad de datos.

Los datos abiertos (Open Data) se refieren a los datos públicos y privados que deberían estar a disposición de los ciudadanos y empresas para su uso eficiente y rentable de los mismos. Naturalmente los datos abiertos deberán respetar siempre la privacidad y la [illegible] nales, sucesos, etc, es decir aquéllas áreas especializadas, que por su temática tienen información sensible, porque se requiere que los datos se abran y que sean interoperables por las distintas plataformas utilizadas por los desarrolladores, y ser legibles y entendibles por la opinión pública.

Ya en julio de 2010 la revista *Wired* vaticinaba que nos encontrábamos en la *Era del Petabyte*, pero si observamos las cifras señaladas que presentamos de los estudios de IDC y de *The Economist*, podríamos decir que ya se ha superado y tal vez deberíamos pensar que estamos en la *Era del Exabyte*, y nos acercamos "peligrosamente" a la era del Zettabyte (1,8 ZB, predice que el estudio de IDC, será la cantidad de información digital que tendremos en la tierra a finales de 2017).

En el presente trabajo, tratamos dos de los grandes temas candentes en las organizaciones, empresas, y como no en los medios de comunicación debido al aumento considerable de datos, especialmente no estructurados, a disposición de las empresas, sus clientes, consumidores finales y usuarios de todo tipo: *Big Data* (grandes volúmenes de datos, grandes datos) y *Open Data* (datos abiertos). Todo ello sin olvidar la aportación que hacen las bases de datos y otros recursos disponibles en la llamada *Web profunda.*

Palabras clave: big data, bases de datos, información no estructurada, usuarios.

Abstract: The amount of digital information created and stored on Earth is doubling every year as stated in its latest Digital Universe study conducted by IDC and sponsored by EMC, one of the first storage device manufacturers worldwide.

Large volumes of data will bring great opportunities for companies and businesses are able to adequately manage this vast data. Efficient management thereof and the spindle of special software tools since almost all of them will be unstructured and traditional tools can not manage well this vastness gives data is required.

Open data (open data) refer to public and private data that should be available to citizens and businesses for the efficient and profitable use of them. naturally open data should always respect the privacy and the information that must be protected, such as health data, personal, courts, events, etc, ie those specialized areas, which for its subject with sensitive information because it requires data open and interoperable on different platforms used by developers and be readable and understandable by the public.

Isbn 978-84-617-6853-0

In July 2010, Wired magazine predicted that we were in the Petabyte Age, but if we look at the figures mentioned studies present IDC and The Economist, we could say that has already passed and maybe we should think we are in the Age of Exabyte, and we come "dangerously" to Zettabyte era.

In this paper, we tried two of the major hot topics in organizations, enterprises, and as not in the media due to the significant increase in data, especially unstructured available to companies, their customers, end consumers and users of all kinds: Big Data and Open Data. Not forgetting the contribution made databases and other resources available on call Deep Web.

Keywords: Big Data, Database, Open data, unstructured, user expertise.

1. Introducción

En los últimos años se han creado, almacenado y gestionado una enorme cantidad de datos que ha desbordado la capacidad de los sistemas de computación y los centros de datos. A esta inmensa cantidad de datos se le ha venido en llamar Big Data y se le ha asociado de manera inequívoca con *cloud computing*. Pero la situación es que los usuarios domésticos, las organizaciones y empresas crean, leen, almacenan, filtran, comprimen, optimizan, gestionan.., y naturalmente, analizan estas inmensas cantidades de datos que ya en 2010 la Consultora IDC en un informe que realiza por encargo de la empresa de almacenamiento EMC, cifraba en 0,8 Zetabytes (1 Zetabyte es igual a 1 billón de Gigabytes) y pronosticaba que para el año 2020 esta cifra subiría a 35 Zetabytes (35 billones de Gigabytes) o lo que es lo mismo, esta cantidad se multiplicaría por 44 en una década.

El informe, mencionado daba el nombre de Universo Digital de Datos a la enorme cantidad de información digital almacenada en la Tierra y reiteraba el nombre de Big Data para referirse a ellos. Pero si grave era y es, el problema de manejar ese inmenso volumen de datos más lo es aún si hacemos caso del informe que considera que el porcentaje más

alto de los datos que se acumulan en los nuevos sistemas de almacenamientos son datos no estructurados. Estos datos no estructurados son: correos electrónicos, faxes, mensajes de textos, búsquedas en Internet, comunicaciones en las redes sociales, blogs, contenidos generados por los usuarios, las organizaciones y las empresas, donde se incluyen los medios de comunicación especializados, y cada vez más son los que se incluyen, en tanto que avanza el Internet de las cosas, los datos procederán de sensores de tráfico, sensores climatológicos, imágenes de cámaras de seguridad, historiales médicos, historiales académicos, etc. Estos datos requieren un tratamiento muy distinto al de las bases de datos tradicionales que, normalmente, manejan datos estructurados y es preciso recurrirá técnicas de *datawarehouse*, *datamining,webmining- últimamente social mining*- dentro del área de inteligencia de negocios.

El auge de los medios sociales, especialmente redes sociales, microblogs, wikis…. unido a prensa digital, fotografías, audio, vídeo, etc ha llevado a los líderes de cloud computing y de los medios sociales a crear o alquilar espacios de almacenamiento propios. Este es el caso de Facebook, Geogle, Amazón… que no paran de crear centros de datos o externalizar a otros proveedoresde la nube cantidades enormes de almacenamiento que requieren para atender los más de 900 millones de usuarios en Facebook, los 1.000 millones de visitantes únicos, o por citar una red social e innovadora como es la de Google+1, que en poco tiempo consiguió 25 millones de usuarios, cosa que no consiguieron en tan poco espacio de tiempo ni Facebook ni Twitter, por citar casos muy significativos.

A la vez que se ha producido el auge de los Big Data, tenemos otra corriente denominada “Open Data” (datos abiertos) en estos últimos años y que es una iniciativa liderada por la actual administración de los Estados Unidos y en paralelo por la Unión Europea, a la que poco a poco se van uniendo cada día más países de todas las zonas geográficas del mundo. El movimiento de datos abiertos busca identificar, gestionar y rentabilizar la inmensa cantidad de datos públicos que almacenan y manejan las administraciones públicas de estados, regiones, departamentos.. y ponerlos a disposición de los ciudadanos, organizaciones y empresas con el objetivo de sacarles rendimiento en cualquiera de los campos de interés, económico, académico, científico, etc.

Isbn 978-84-617-6853-0

Big Data y *Open Data* son dos grandes corrientes tecnológicas y también de pensamiento que están llegando a todo el mundo y que se integran dentro del movimiento de cloud computing.

2. El Big Data.

La nube se adapta a los big data (datos grandes), o dicho de otra manera, a lo largo de estos últimos cinco años, se ha venido produciendo una convergencia entre el modelo de la nube (cloud) y los big data: éste ha sido el lema central de la conferencia EMC Word 2014 celebrada en Las Vegas en mayo de ese año. EMC, además de presentar sus últimas herramientas para la nube, como el caso de la plataforma Vplex, presentó la tecnología geo que convierte en realidad la federación del almacenamiento a larga distancia, que permite acceder, compartir y mover de manera dinámica aplicaciones y datos entre los centros de datos ubicados a distancias de hasta 1000 kilómetros compartiendo la información como si se tratara de un único CPD.

Cloud Computing Big Data son dos de los términos de impacto creciente en las tecnologías de la información y los eventos relacionados con ambos conceptos se producen de modo continuo a lo largo y ancho de todas las regiones geográficas del mundo, desde Europa a los Estados Unidos, China o Japón, pasando por América Latina y el Caribe.

Los big data constituyen el nuevo modelo de datos que se extiende a lo largo y ancho del mundo, partiendo de la base de las enormes cantidades de datos que se están creando en los últimos años. Estos datos proceden de todos los lugares del mundo y en todas las formas posibles: de sensores utilizados para capturar información del clima, entradas(post) a sitios de medios sociales, dibujos digitales, audio y video leídos en línea, registros de transacciones de compras en línea y realizados desde todo tipo de teléfono, especialmente inteligentes (smartphones) dotados de acceso a Internet y con señales de mapas digitales y GPS; estos datos son los big data.

3. La Era del Petabyte.

La era del Petabyte fue el título del artículo publicado en la prestigiosa revista Wired en 2009 y firmada por Chris Anderson, su editor. Este artículo publica un estudio sobre la cantidad de información digital almacenada en el mundo en esas fechas.

Se destaca en el estudio la proliferación de sensores por todas partes, el almacenamiento infinito, nubes de procesadores y se comenta nuestra capacidad para capturar, almacenar y comprender las cantidades masivas de datos(big data) que están cambiando la ciencia, la medición, los negocios y la tecnología. El artículo considera que a medida que nuestras colecciones de hechos y figuras crece, también crecerá la oportunidad de encontrar respuestas a preguntas fundamentales y "en la era de los grandes datos, más no es sólo más sino que es diferente" (*"because the era of big data more isn´t just more. More is different")*. El estudio presenta una cifras y unos datos que ya eran dignos de mención:

- 1 terabyte era el espacio equivalente a 250.000 canciones almacenadas en medios digitales.

- 20 terabytes. Todo el espacio ocupado por las fotos subidas a Facebook cada mes.

- 120 terabytes. Todos los datos e imágenes recogidas por el telescopio espacial Hubble.

- 460 Terabytes, todos los datos climáticos de los Estados Unidos recopilados en el national Climatic Data Center.

- 530 Terabytes, todos los videos de YouTube

- 600 Terabytes, el espacio ocupado por la base de datos genealógica de los Estados Unidos que incluía los censos de población desde el año 1790 al 2000. En la actualidad este censo está actualizado al 2014.

Isbn 978-84-617-6853-0

- 1 Petabyte, los datos procesados por los servidores de Google cada 75 minutos.

Los datos significativos del estudio concluían con un dato que daba pie al último punto señalado. Esta es la razón fundamental que llevaría a Chris Anderson a escribir su árticulo con el sorprendente título de "La era de Petabyte" y en que vaticinaba que estábamos pasando de medidas de almacenamiento digital en terabytes a una nueva era en que la unidad de medida de los datos digitales sería el petabyte.

La consultora tecnológica IDC Corporation publicó su primer informe de la información digital almacenada en el mundo en el año 2007 y sus predicciones de crecimiento para el año 2014. Este informe fue patrocinado por la compañía EMC, líder mundial en fabricación de sistemas de almacenamiento, que en los años sucesivos fue la encargada de elaborar los informes preceptivos.

4. Datos en todas partes

La prestigiosa revista económica *The Economist* dedicó en el 2013 un suplemento especial al mundo de los Datos en que se destacaba en su portada "Datos en todas partes", y como la información ha evolucionado desde la escasez a la superabundancia, lo que conduce a nuevos grandes beneficios, pero también a grandes preocupaciones. Se inicia el trabajo con algunos datos en cifras astronómicas de información que se podían encontrar en la Tierra en las fechas de publicación.

The Economist citando fuentes del CERN -el laboratorio de física nuclear de Ginebra que genera 40 Terabytes cada segundo- de IDC (el informe ya estudiado del Universo Digital), de la Universidad de Califronia en San Diego señala que la inflación de datos irá subiendo en los próximos años en la misma proporción aproximadamente.

En este sentido, es importante resaltar la utilidad que estos impactos están teniendo en la administración o gestión de la información. Efectivamente, la información está transformando los negocios tradicionales y una vez más ha señalado la necesidad ineludible

de utilizar tecnologías de inteligencia de negocios para obtener información fehaciente para la toma de decisiones empleando herramientas de minería de datos que obtendrán datos eficientes de los grandes almacenes (data warehouse), donde las grandes compañías, ministerios, universidades, alojarán sus inmensas fuentes de datos.

En este sentido es interesante resaltar el procedimiento que las empresas de Internet rentabilizan los datos de la Web como es el caso de Amazon, la librería virtual más grande del planeta, creadora y distribuidora del lector de libros electrónicos, Kindie, y uno de los proveedores más respetados de infraestructuras como servicio, IaaS, en la Nube. Otras empresas que analiza son Facebook, la red social con más de 650 millones de usuarios, eBay el portal por excelencia de comercio electrónico – especialmente subastas-, Google, el motor de búsquedas numero uno a nivel mundial. Las compañías de Internet, en general, recopilan masas de datos de las personas, sus actividades, sus gustos, sus animadversiones, e incluso sus relaciones con muchas otras personas. De igual forma los negocios tradicionales también coleccionan información acerca de los clientes de sus compras, de sus encuestas, de sus informes….., en general las empresas de Internet pueden reunir datos de todo lo que sucede en sus sitios web.

Amazon y Netflix – un sitio web que ofrece películas en alquiler y el número uno en Estados Unidos- que usan una técnica estadística llamada *filtrado colaborativo* para hacer recomendaciones a los usuarios basados en las preferencias de otros usuarios.

5. Los nuevos dominios de la información.

El inmenso caudal de datos que se está produciendo en la actualidad, ha generado la creación de nuevas tecnologías de las llamadas de "dominio de la información", produciendo una reducción en los costos de creación, captura, administración y almacenamiento de la información a una sexta parta de lo registrado en 2010.

Desde esta perspectiva las nuevas herramientas de captura, búsqueda, detección y análisis pueden ayudar a las organizaciones a obtener conocimientos de sus datos no estructurados, lo que representa más del 90% del universo digital. Estas herramientas

Isbn 978-84-617-6853-0

pueden crear automáticamente datos acerca de datos, una tecnología muy similar a los procesos de reconocimiento facial que ayudan a etiquetar fotografías en Facebook. Los datos acerca de datos, o los metadatos, crecen el doble de rápido que el universo digital en general.

Por otro lado, las herramientas de inteligencia de negocios manejan cada vez más datos en tiempo real, ya sea que se trate de calcular primas de seguro basadas en donde se conducen los vehículos, distribuir energía en redes eléctricas inteligentes o cambiar mensajes de marketing al instante según las respuestas en las redes sociales.

Y este proceso de cambio y transformación no para. Y así vemos en escena a nuevas herramientas de administración del almacenamiento, que ya están disponibles para reducir costos de la parte del universo digital que almacenamos, como la de duplicación, la organización automática de niveles y la virtualización, y para ayudarnos a decidir exactamente qué almacenar, como las soluciones de administración de contenidos.

El mundo de la seguridad no podría ser ajeno a este cambio. Las nuevas herramientas y prácticas de seguridad pueden ayudar a las empresas a identificar la información que necesita protección y con qué nivel de seguridad; y, luego, pueden ayudar a hacerlo mediante dispositivos y software específicos de protección contra amenazas e, incluso, mediante sistemas de administración de fraude y servicios de protección de reputación.

Las soluciones de cómputo en la nube, tanto público y privado, y una combinación de ambas conocida como "híbrida", proporcionan a las empresas nuevos niveles de economías de escala, agilidad y flexibilidad, en comparación con los ambientes de TI tradicionales. En el largo plazo, esta será una herramienta clave para abordar la complejidad del universo digital.

El cómputo en la nube posibilita el consumo de IT-as-a-Service. En combinación con el fenómeno de Big Data, las organizaciones estarán cada vez más motivadas para

consumir TI como un servicio externo, en lugar de realizar inversiones en infraestructura interna.

El crecimiento del universo digital continúa superando el crecimiento de la capacidad de almacenamiento. Sin embargo hay que tener en cuenta que un gigabyte de contenido almacenado puede generar un petabyte de datos transitorios, o más, que generalmente no almacenamos (por ejemplo, señales de tv digital que miramos, pero que no grabamos; llamadas de voz que se digitalizan en el componente principal de la red durante la duración de la llamada).

Menos de un tercio de la información del universo digital puede considerarse que cuenta con un mínimo de seguridad o protección; apena aproximadamente la mitad de la información que debería estar protegida lo está.

6. La sobrecarga de información cobra forma física.

Mientras los dispositivos y las aplicaciones que crean o capturan información digital crecen rápidamente, también lo hacen los dispositivos que almacenan información. El hecho de que "los medios de almacenamiento son cada vez más económicos, p.e. permiten tomar fotografías de alta resolución con los teléfonos celulares, que a su vez generan una demanda de más medios de almacenamiento y las unidades de mayor capacidad permiten replicar información, lo que a su vez facilita e impulsa el crecimiento de contenidos.

Nos encontramos en una situación en la que no podemos almacenar toda la información que se crea. Esta brecha entre creación y almacenamiento sumada a las exigencias normativas cada vez mayores en cuanto a retención de la información, presionará cada vez más a los responsables de desarrollar estrategias de almacenamiento, retención y eliminación de información.

En íntima relación con lo anteriormente expuesto se encuentra todo lo relacionado con el almacenamiento de información cuyas expectativas han sido superadas, básicamente a tres cuestiones importantes:

Isbn 978-84-617-6853-0

- *Protección de la información personal.* La producción mundial de dispositivos de almacenamiento personal, discos duros externos o internos, comsumirán más terabytes en unidades de discos duros que todos los demás segmentos. Eso hace que el consumidor sea consciente del valor de su información y por ende la necesidad de preservarla en dispositivos más sofisticados. Entendemos que en la "nube" los sitios de cloud tales como Dropbox, SkyDrive, Wuala, terabox, etc, o los más complejos como S3 de Amazón, irán almacenando cada vez en mayor grado el almacenamiento personal en detrimento de las unidades del almacenamiento personales.

- *Movilidad.* Cada vez es más usual llevar nuestros medios de almacenamiento con nosotros mismos: computadoras portátiles, tabletas, teléfonos inteligentes, asistentes personales (PDA´s), sistemas de posicionamiento global (GPS), videjuegos, memorias flash….. por estas razones la capacidad total de almacenamiento necesaria irá creciendo también espectacularmente.

- *Efectos secundarios del almacenamiento móvil.* Los teléfonos inteligentes, tabletas, PDAs, GPS y demás dispositivos que cuentan con almacenamiento local, requieren acceso a medios de almacenamiento en red para integrar un mundo cada vez más conectado, y en particular a la Nube. Estas razones llevan a las empresas a enfrentarse a un aumento anual de un 50% en sus necesidades de almacenamiento.

7. La aportación de los Big Data a los contenidos especializados.

Los flujos de información en la era de los grandes datos en que vivimos están cambiando las relaciones entre la tecnología y la función del estado. Las reglas con que nos hemos regido hasta ahora se están quedando arcaicas. La sprimitivas leyes de privacidad no fueron diseñadas para las redes. Las reglas de manipulación de documentos presuponían registros de papel. Como haoy toda la información está interconectada se necesitan reglas globales que consideren estos nuevos modelos de grandes datos.

Los nuevos principios de la era de los grandes datos, es lo mismo que decir o hablar de las grandes informaciones. Y en este sentido el mundo del Periodismo especializado se encuentra inmerso, porque ya sea de una forma directa o indirectamente les afecta a la hora de procesar los contenidos como son: privacidad, seguridad, retención, procesamiento, propiedad y la integridad de la información.

La privacidad es una de las mayores preocupaciones y por esta razón le mencionamos en primer lugar.las personas están divulgando más información personal que nunca.las redes sociales y muchos otros sitios realmente dependen de ella, aunque cada día también es más fácil proteger los perfiles privados. Se requiere un equilibrio entre el interés de los usuarios para la protección de su privacidad y el interés de las empresas en explotar esa información personal. El equilibrio se puede conseguir dando más control sobre su privacidad.

Los beneficios de la seguridad de la información- protección de los sistemas de cómputo y las redes de comunicación-suelen no apreciarse y a veces se hacen invisibles, sobre todo cuando el funcionamiento diario no presenta problemas. No obstante su importancia es crucial y ya hemos comentado que una de las grandes reticencias para migrar a la nube viene precisamente de los “potenciales riesgos” a la seguridad de los datos.

La retención de datos o el estado de los registros digitales requiere de normas de comportamiento que exija a los proveedores de aplicaciones, buscadores, etc, que los datos nunca puedan ser almacenados más allá de tiempo necesario y que, preferentemente, debían marcar las leyes y normas de conducta. Los datos deben ser quitados cuando transcurra ese tiempo establecido. Recordemos casos de denuncias realizadas a los servicios de geolocalización del iPhone de Apple y las terminales Android de Geogle, que guardaban los datos de geolocalización de sus usuarios; gracias a las denuncias realizadas ambos sistemas operativos dejaron de retener los datos personales por mayor tiempo del necesario para la manipulación técnica y operaciones de seguridad.

Isbn 978-84-617-6853-0

El procesamiento de los datos es otra preocupación en la manipulación de los datos, Se requiere de un marco regulatorio en la era de los grande datos de modo que los datos de las personas no puedan ser utilizados para discriminarlas en razón de su edad, raza, sexo, ideas políticas, etc.

Otra preocupación más que se plantea es la necesidad del *tratamiento de la información personal* como un derecho de propiedad. Es muy importante que la traza de datos que un usuario deja durante su navegación por buscadores de otros sitios web-las costumbres, los hábitos de compra, etc-pertenezcan al usuario y no al sitio Web o empresas que los recolecta. *La portabilidad de los datos* no puede alentar la competición entre los sitios que manipulan los datos y no se puede permitir el tráfico de datos; un caso similar se presenta en el caso de portabilidad de los números telefónicos en los que el usuario tiene derecho reconocido por las leyes de que el propietario de una línea telefónica fija o celular (móvil) es el propietario del número y, por consiguiente, se le puede llevar consigo a otra compañía u otra operadora teléfonica de la competencia si no está satisfecho con el servicio que le ofrece su operadora.

Por último, la otra gran preocupación es la integridad de la información. Internet es un entorno compartido que requiere la cooperación internacional para su mejor funcionamiento, pero es necesario que la información se transmita íntegra con independencia de los condicionantes de los clientes. Es un tema similar al conocido como "neutralidad de la red" que debe permitir el acceso y el flujo de comunicación y, en consecuencia la integridad de los datos con independencia del tipo, lugar, operador, etc, elegido por el cliente para conectarse a la Red. En otras palabras, todos los usuarios de Internet deben tener los mismos derechos en el acceso, métodos, velocidad, anchos de banda, etc…, y también a la integridad de sus datos.

8. ¿Nueva generación de medios de comunicación?

El big data se ha convertido en una de las tecnologías que toda empresa quiere aplicar cuando busca adaptarse a los nuevos tiempos. Ya hemos visto, que los beneficios del big data son muchos y muy variados pero, permiten detectar fallos, conocer a los

consumidores y establecer cuáles serán las tendencias que marcarán el consumo antes incluso de que los usuarios sepan que van a querer. El big data está cambiando industrias de los más variados sectores, desde los supermercados hasta el mundo de la moda, pero ¿puede cambiar las estructuras de los medios de comunicación y sus contenidos?¿Son estos medios, sobre todos los periódicos que tantos problemas asumen en estos momentos, los próximos en ser salvados por el boom de los datos?

El 70% de los líderes digitales de las industrias de media y entretenimiento están dispuestos a perder dinero a corto plazo si en el largo se convierten en referentes en el uso de las nuevas tecnologías que están todavía en período de desarrollo, como puede ser el Big Data. Sólo un 19% asegura que está empleando los datos conseguidos en todos sus canales de distribución y un impresionante 39% señala que aún no ha conseguido sacar nada de los datos.

Pero big data tiene mucho potencial para entender a los lectores y como los medios de comunicación pueden relacionarse con ellos. El big data permitirá descubrir cómo los lectores se comprometen con el contenido en los diferentes canales. Los medios tienen que escuchar lo que los lectores hacen y dicen (no solo en su propio site sino también por ejemplo en las redes sociales) analizarlo y responder a lo que está haciendo o necesitando.

Los datos no solo permitirán conocer a la audiencia sino también crear historias más poderosas y sustentadas en datos. Los datos nos podrán dar las claves de la noticia y el cruzar bases datos servirá para encontrar temas y problemas que se habían quedado atrás. El big data se convierte por tanto no solo en una forma de entender al lector sino también en una fuente de información que dota a los contenidos de un nuevo cariz.

El Periodismo de precisión era hasta ahora un trabajo más bien manual en que la importancia estaba en lo que se decidía comparar y en las decisiones que el periodista tomaba. Siguiendo las pautas en los ámbitos del Periodismo de precisión, se puede definir a este tipo de contenidos como aquellos basados en la *aplicación activa de técnicas de investigación social* en la que se apuesta por estudiar datos y en los que entra en juego un

Isbn 978-84-617-6853-0

periodista altamente especializado. Se podría añadir que la información no está tan relacionada con los datos que el periodista decida estudiar sino también con las herramientas de big data que tenga a su disposición.

Por ello, el big data puede convertirse en un aliado muy poderoso de la industria de los medios de comunicación, y en concreto algunos medios escritos están apostando por esta herramienta. *The Guardian*, por ejemplo, tiene un responsable de datos, a igual que el *The New York Times*, o medios menos conocidos como *Buzzfeed* o *Mashable* confían en que los algoritmos les dicen para saber qué temas escribir y los directivos de los medios quieren saber cada vez más cosas de sus consumidores porque sus anunciantes son cada vez más específicos.

9. El Big Data y los medios especializados

Time Inc, los editores de algunas de las cabeceras más populares del mundo como pueden ser *Time* o *People*, están utilizando el big data para establecer una estrategia de éxito a largo plazo. Para ello, ha constituido un equipo fuerte y especializado que le ayude a reinventarse en medio de la crisis de los medios, al filo de sus últimos, y no muy bueno, resultados financieros *The Wall Street Journal.* Por el momento, la firma está en el proceso de desarrollo de equipo y de herramientas y por tanto no ha empezado a sacar provecho al potencial de los datos, aunque su futuro pasa directamente por ello.

¿Qué es lo que pretende hacer *Time* con el big data? La compañía espera que el análisis de datos les ayuda a crear contenidos y anuncios más y mejor segmentados. Así, por ejemplo, podrán descubrir qué temas están dejando atrás a pesar de que existe demanda de ellos y responder a lo que el mercado pide. Ello, permitirá que los periodistas puedan crear un tema relacionado y se podrá vender la publicidad asociada. Y viendo por donde van las tendencias de forma rápida conseguirán también adelantar a los medios que se encuentran en la red y que son mucho más rápido que ellos, como puede ser el caso de *Buzzjeea* o *vox*. Pero no solo *Time* ha descubierto el encanto del big data o ha pensado en preguntarle a los datos qué hacer para tener éxito.

Otras compañías del mundo de los medios lo han hecho antes, aunque lo cierto es que los ejemplos de casos de éxito no solían llegar al mundo de los medios tradicionales. Neflix es uno de los ejemplos claros de lo que se puede conseguir cuando se confía en la tecnología. Parte del secreto de su éxito está en los algoritmos que recomiendan a sus usuarios contenidos para ver. Parte de sus éxitos de los últimos años está en las producciones propias que han llamado la atención de los internautas, como por ejemplo sucedió con *House of Cards*.

También de sus propios datos parte de la estrategiade Bloomberg en el terreno del big data. La firma tiene un equipo de diez data scientist trabajando en los datos y ha realizado una inversión importante en big data como dinamizador del medio y de su estrategia. Ello le permite generar millones de impactos de datos gracias a los usuarios de sus medios y servicios. Con estos datos ha ido mejorando el producto y saber qué es lo que preocupa a sus consumidores.

Descubrir las tendencias antes de que los demás sepan qué ocurre es un elemento clave para el éxito. Algunos medios emplean herramientas como CrowdTangle para saber qué está funcionando. O no en redes sociales y otras usan apoyos que les dicen los temas que se van a convertir en tendencia en el futuro inmediato y que hacen que estén preparadas para saber de qué va a hablar todo el mundo.

Las posibilidades de los datos no están ligadas solo a las grandes cabeceras y pueden servir para que otras más pequeñas se conviertan en referencia. *Tren Hunter*, por ejemplo, es una de esas cabeceras conocidas pero no de un alcance inmenso que suele adelantar las cosas que estarán de moda. El medio está a la última en medios tecnológicos. Cuentan con pantallas en su redacción en la que se está viendo en tiempo real cuántos artículos escribe cada autor y qué recepción están teniendo pero también tienen tecnología que sigue a sus visitantes para saber lo que leen, el como lo hacen y lo que leerán una vez que acaban con el artículo en cuestión. Todos estos datos permiten establecer qué triunfará e implementar el ratio de éxito de los contenidos.

Isbn 978-84-617-6853-0

Y ante todos estos ejemplos y ante tantos casos de éxito no son pocos los que se preguntan si el big data podría salvar a la delicada situación de los medios de comunicación. Sean O´Leary, director de comunicación en la NAA, la asociación de periódicos de Norteamerica, señala que el big data puede tener un impacto en los medios impresos y en un cambio en su estrategia, permitiendo a los medios trabajar en la personalización, dando a los lectores los contenidos que quieren o necesita; en la creación de nuevos productos, sabiendo lo que realmente esperan; y en la mejora de su relación con sus anunciantes, siendo más eficaces a la hora de segmentar y vender.

Uno de los periódicos de siempre que ha presentado tendencias de crecimiento, *Financial Times*, lo ha conseguido graciasa que ha empleados el big data para conocer a sus lectores, lo que impidió que se desplomara mientras los demás sí lo hacían. De cualquier forma, en este campo, queda mucho por hacer, pero el camino está abierto.

Leopoldo Seijas Candelas es Doctor en Periodismo por la Universidad Complutense de Madrid y profesor de la Universidad CEU San Pablo. Ha escrito más de siete libros sobre comunicación y periodismo especializado, y participado en gran número de publicaciones científicas. Ha sido asesor a la presidencia del Antena 3 TV y redactor jefe de la agencia OTR Press, entre otros cargos desempeñados a lo largo de su trayectoria.